Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence

Joseph T. McGuire and **Joseph W. Kable**

Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104, USA

Joseph T. McGuire: mcguirej@psych.upenn.edu; Joseph W. Kable: kable@psych.upenn.edu

Abstract

Deciding how long to keep waiting for future rewards is a nontrivial problem, especially when the timing of rewards is uncertain. We report an experiment in which human decision makers waited for rewards in two environments, in which reward-timing statistics favored either a greater or lesser degree of behavioral persistence. We found that decision makers adaptively calibrated their level of persistence for each environment. Functional neuroimaging revealed signals that evolved differently during physically identical delays in the two environments, consistent with a dynamic and context-sensitive reappraisal of subjective value. This effect was observed in a region of ventromedial prefrontal cortex that is sensitive to subjective value in other contexts, demonstrating continuity between valuation mechanisms involved in discrete choice and in temporally extended decisions analogous to foraging. Our findings support a model in which voluntary persistence emerges from dynamic cost/benefit evaluation rather than from a control process that overrides valuation mechanisms.

Pursuing long-run rewards often requires persistence in the face of delay and short-run costs. The capacity to delay gratification is central to the notion of *self-control* in human decision making, and failures of persistence can appear to reflect impulsivity, inconsistency, or self-control failure[1, 2]. Here we used fMRI to examine brain activity associated with sustaining or curtailing persistence toward delayed rewards.

Much is known about neural systems involved in value-based decision making[3–6], but it is unknown what role these mechanisms play in temporally extended persistence. Most intertemporal choice research focuses on discrete choices among outcomes that differ in delay[7–9]. Delay-of-gratification scenarios, in contrast, involve a prolonged delay period with a continuously available opportunity to give up[1]. These two types of future-oriented behavior are widely thought to involve different mental processes. Mischel and colleagues[10]

have argued that the initial selection of a delayed reward depends on a rational cost/benefit assessment, but that the subsequent ability to wait for it depends on self-regulatory dynamics (competition between hot and cool motivational systems[11]).

We previously hypothesized that both successes and apparent failures of persistence emerge from dynamic value maximization[12, 13]. Because the exact timing of future events is usually uncertain, there is no guarantee that a decision maker who was willing to begin waiting for a delayed reward should necessarily be willing to keep waiting indefinitely. In some situations, including many that seem to challenge self-control, a long delay so far is predictive of a longer-than-expected delay yet to come[12–15]. One way to navigate such situations would be to reassess the subjective value of the awaited reward as time passes, based on a continuously updated estimate of the remaining delay time[12]. Such a reassessment might be encoded in the same neural *valuation system*, comprised of ventromedial prefrontal cortex (VMPFC), ventral striatum (VS) and posterior cingulate cortex (PCC), that encodes subjective value in a highly general manner across many other kinds of decisions[3–6]. The subjective value representations encoded in VMPFC are known to be sensitive to both immediate and delayed outcomes[7, 8], primary and secondary forms of reward[3, 16], goal-related and temptation-related factors[9, 17], and high-level task contingencies[18, 19].

Other theoretical perspectives make different predictions. One alternative possibility is that successful persistence depends principally on cognitive control mechanisms external to the valuation system. Although some accounts hold that the VMPFC valuation system mediates cognitive control[9, 17, 20], other accounts posit a form of control that overrides or competes with valuation[2, 11, 20–23]. If the latter control mechanism is paramount, successful persistence might be better understood as rule-adherence than as value-maximization, and curtailing persistence might reflect a lapse in control-related brain activity (e.g., in lateral PFC).

A second alternative possibility is based on the structural parallel between delay of gratification and certain kinds of foraging scenarios[13, 15, 24–27]. It has recently been hypothesized that single-alternative foraging decisions—e.g., whether to exploit one's current food patch or depart to forage elsewhere—might depend, not on the VMPFC valuation system, but on a representation in dorsal anterior cingulate cortex (dACC) of the value of departing[26, 28].

To examine valuation signals during temporally extended persistence we conducted an fMRI experiment in which participants repeatedly decided how long to keep waiting for future monetary rewards (Fig. 1a). On each trial the participant viewed a token, which had no initial value but matured to a value of 30¢ after a random delay. The participant could sell the token anytime and initiate a new trial, aiming to maximize total earnings in a fixed time period. Unlike some previous studies[1, 13], no small reward was delivered if the participant quit early; instead, the main incentive to quit was the possibility that the next trial might mature with a shorter delay.

The ideal strategy depended on the distribution of delay times, which differed between two environments (Fig. 1b,c). In a high-persistence (HP) environment the most productive strategy was to wait for every reward (up to 40s). In a limited-persistence (LP) environment the best strategy was to wait 20s and then quit if the reward had not arrived. Participants learned about the timing statistics through direct experience during preliminary training. The environments were presented in alternating 10-min runs, marked by different-colored tokens.

We predicted that participants would quit earlier in the LP environment than in the HP environment[13]. In addition, our theoretical model predicted that participants' subjective valuation of the awaited token would evolve differently in the two environments, increasing more rapidly with elapsed time in the HP environment than the LP environment. Our neuroimaging experiment tested whether canonically value-responsive brain regions would reflect this dynamic reassessment. Our experiment could also detect alternative possibilities such as representations of subjective value elsewhere in the brain, a lapse in control-related activity associated with quitting, or a representation of the value of quitting in dACC.

Results

Behavioral results

Participants (*n*=20) quit before receiving the reward more often in the LP environment (median=50.0% of trials; IQR 46.6 to 57.6%) than the HP environment (median=3.1%; IQR 0 to 15.6%). In the LP environment the time waited before quitting (median of medians) was 29.3s (IQR 17.6 to 36.6). Within-subject (across-trial) variability in quit timing was comparatively small: the median size of the within-subject interquartile range was 9.1s.

Participants were willing to wait longer in the HP environment than the LP environment. We used survival analysis to estimate each participant's probability of "surviving" various lengths of time without quitting[13]. Fig. 2a shows averaged subject-wise empirical survival curves against ideal performance. The area under the curve (AUC) estimates how much of the first 40s a participant was willing to wait on average (Fig. 2b). Median AUC was 38.9s in the HP environment (IQR 35.4 to 40; ideal=40s) and 30.2s in the LP environment (IQR 22.3 to 34.9; ideal=20s). All 20 participants persisted longer in the HP environment (median difference=7.6s, IQR 3.0 to 14.2, signed-rank *p*<0.001). Persistence in the two environments was modestly correlated (Spearman $\rho_{n=20}$=0.37, *p*=0.11; Fig. 2b), and behavior was stable across the fMRI experiment (Supplementary Fig. 1).

Reaction time (RT) to sell rewarded tokens tracked time-varying reward expectancy. When an event's latency is uniformly distributed, expectancy theoretically increases with elapsed time[28] (Fig. 2c). Accordingly, subject-wise Spearman correlations between delay and RT were reliably negative in the HP environment (median single-subject ρ=−0.27, IQR −0.36 to −0.16, signed-rank *p*<0.001), indicating faster responses to rewards that were preceded by longer delays (Fig. 2d) and implying that participants successfully encoded the task's timing statistics.

Theoretical modeling

The passage of time can drive a dynamic reassessment of awaited rewards by furnishing information about the remaining delay[12, 29]. Intuitively, rewards in the HP environment grew nearer and more subjectively valuable as time passed, but rewards in the LP environment became progressively less likely to be delivered before the participant quit.

We formalized this intuition in a theoretical model of subjective valuation. The model estimated the awaited token's subjective value at each point in the delay interval, accounting for the changing probability distribution over remaining delay durations. Our model extended a formalism from the optimal foraging literature known as the *potential function*[25]. The expected remaining delay was multiplied by the opportunity cost of time and subtracted from the expected reward. Subjective value at a given elapsed time equaled the expected net return in the remainder of the current trial, maximized over all possible giving-up times. Its minimum was zero since the agent could always quit immediately. If subjective value exceeded zero, this signified that the decision maker could do better by waiting than by quitting immediately. The level of subjective value at each time reflected the margin of preference for waiting over quitting (see *Methods* for details).

In the HP environment the token's theoretical subjective value increased with elapsed time, reflecting the progressive shortening of the expected remaining delay (Fig. 3a). In the LP environment the token's subjective value remained positive until 20s but then fell to zero, reflecting that the best strategy was to quit if the reward had not arrived by then. Differences between the subjective value trajectories in the two environments were primarily driven by the evolving probability that the reward would arrive before the optimal giving-up time (Supplementary Fig. 2).

We modeled the empirical behavioral data as a stochastic function of theoretical subjective value using logistic regression (Fig. 3b; see *Methods* for details). Greater subjective value was associated with higher odds of waiting (median coefficient=0.26, IQR 0.05 to 0.78, signed-rank $p<0.001$). The subjective value model significantly outperformed an intercept-only model (subject-wise likelihood-ratio tests: median z=4.26, IQR 1.79 to 7.97, signed-rank $p<0.001$) and an alternative model that directly fit different overall rates of quitting in the HP and LP conditions (subject-wise difference of model deviances: median=6.45, IQR −1.85 to 32.02, signed-rank p=0.033).

Neuroimaging results

Our fMRI analyses tested for brain signals that evolved differently during physically identical delay intervals in the two environments. Trial-onset-locked BOLD timecourses were flexibly estimated in each environment using a finite impulse response (FIR) model; i.e., a series of single-timepoint basis functions in a general linear model (GLM). Each trial was modeled from onset up to 1s before the outcome (reward cue or quit response). Because trials had different durations, earlier timepoints were observed on more trials than later timepoints (Supplementary Fig. 3). Group analyses focused on the interval from 2.5–30s, for which 19 of 20 participants contributed complete data. Because the HP and LP conditions were presented in separate scanning runs with independent baselines our analyses focused

on differential change over time, not the overall offset between the two conditions. Significance was assessed using whole-brain permutation tests to control for multiple comparisons (see *Methods*).

A model-based fMRI contrast tested directly for effects of theoretical subjective value on BOLD activity. For each subject and voxel, the empirical difference timecourse (HP minus LP) was regressed on the predicted difference (Fig. 3c,d; see *Methods*) and a constant intercept. The resulting contrast coefficient reflected the degree to which BOLD signal increased more steeply with elapsed time in the HP environment than the LP environment. Coefficients were submitted to a whole-brain, two-tailed, group-level test (*n*=20). This identified a single significant cluster, located in VMPFC (Fig. 4a and Table 1a), in which BOLD signal was positively related to theoretical subjective value. No negative effects of subjective value on BOLD were identified, even in follow-up analyses tailored to detect signals reflecting the difficulty of persistence (Supplementary Fig. 4).

The observed effect in VMPFC echoes effects of subjective value that are seen in a broad range of other contexts[3–6]. We formally juxtaposed our results with previous findings by quantifying the spatial overlap between our empirical results and canonically valuation-related brain regions derived from a 206-study meta-analysis[3] (Fig. 4a). The meta-analysis had identified clusters showing preferentially positive effects of value in VMPFC ($9.67cm^3$), striatum ($21.41cm^3$), and PCC ($2.62cm^3$). There was a 100-voxel ($2.70cm^3$) region of overlap in VMPFC (27.9% of the canonical region and 32.2% of the empirical cluster).

As an alternative test of the same question, the three canonical valuation areas were tested as regions of interest (ROIs). Model-based contrast coefficients were spatially averaged in each ROI for each participant. The effect of subjective value was significant in VMPFC (signed-rank *p*=0.002) but non-significant, albeit with a positive trend, in striatum (*p*=0.079) and PCC (*p*=0.062; Fig. 4b). Paired-samples comparisons identified a greater effect in VMPFC than striatum (signed-rank *p*=0.012) and no significant differences between the other two pairs of ROIs (*p*s>0.11).

In summary, results suggested that the region of VMPFC previously found to encode subjective value during discrete choices and outcomes also reflected a dynamic reassessment of subjective value during voluntary persistence. This was true to a greater degree for VMPFC than striatum.

We additionally conducted a less-constrained analysis that could detect BOLD timecourse differences predicted either by our model or alternative frameworks. Trial-onset-locked timecourses were analyzed at the group level in a whole-brain voxelwise repeated-measures ANOVA (*n*=19), with factors for condition (HP vs. LP) and timepoint. We focused on the condition-by-timepoint interaction, seeking to identify signals that exhibited different patterns of change over time in the two environments. This analysis avoids *a priori* assumptions about either the form of the difference or the location of effects in the brain. A significant interaction was observed in left and right VMPFC, left posterior parietal cortex, and a small region of left superior temporal gyrus (Table 1b and Fig. 5a). Timecourse plots (Fig. 5b – e) suggested that in VMPFC and parietal regions the effect took the form of a

greater signal increase with elapsed time in the HP environment, consistent with theoretical subjective value.

Further analyses tested for evidence of reward prediction error (RPE) signals[30]. When a reward occurs, RPE is the difference between the obtained and expected outcome. Because reward expectancy theoretically rose over time in the HP environment (Fig. 2c; see also RT data above and heart rate data below), rewards at short delays should have been more surprising and evoked larger RPEs than rewards at longer delays. We tested whether the amplitude of the phasic BOLD response to reward was modulated by the delay duration that preceded it. A negative effect would reflect an RPE-like pattern (smaller reward responses after longer delays, a pattern seen previously in the firing rates of dopaminergic midbrain neurons[31]). To focus on phasic reward responses while controlling for nonspecific effects of elapsed time, we compared the modulatory effect in the post-reward epoch against the same effect in a pre-reward epoch.

We observed no significant negative modulatory effect of elapsed time on the reward-related BOLD response in any location. We did, however, identify an occipitoparietal cluster with an effect in the opposite direction: a higher-amplitude BOLD response to rewards at longer delays, which theoretically were more strongly anticipated (Supplementary Fig. 5). Expectancy-driven amplification of brain responses has been seen before[32], including in visual cortex[33]; these effects bear a family resemblance to the facilitatory effects of spatial attention[34].

Numerous brain areas responded differentially to reward and quit keypresses, including some that exhibited a ramp-up in activity prior to quit responses. We used a GLM to estimate subject-wise perievent timecourses for the two event types separately (using all keypresses across all four runs), and submitted the difference between reward-related and quit-related timecourses to a group-level ANOVA. Significant effects occurred diffusely across DMFC, lateral PFC, anterior insula, precentral sulcus, and occipital and posterior parietal cortex (Fig. 6a – f). In DMFC, anterior insula, posterior parietal cortex, and anterior PFC the difference consisted of an earlier elevated response for quit responses than reward responses. Other regions, including occipital cortex and left inferior frontal gyrus (IFG), responded more strongly to rewards. Broadly, these effects reflect that rewards involved a visual cue whereas quitting was freely timed and volitional.

To test directly for signal changes that preceded decisions to quit, we performed a group-level ANOVA on only the first 5 points in the quit-related timecourse (−12.5 to −2.5s). A significant effect of timepoint within this anticipatory interval was observed in posterior parietal cortex, DMFC, anterior insula, and anterior PFC (Fig. 6a – f and Table 1c). VMPFC showed no effects in either of the above analyses; that is, there was no evidence that subjective value effects in VMPFC could be alternatively explained in terms of a role in response preparation.

Somatic arousal

To test whether subjective value effects in BOLD activity were accompanied by changes in general physiological arousal, we performed exploratory analyses of heart rate (inter-beat

interval measured via pulse oximetry; *n*=17) as a function of task events. Heart rate transiently accelerated after keypresses, but did not differ between the two conditions as a function of delay time (Fig. 7a). In the HP condition there was greater transient cardiac acceleration for rewards preceded by longer delays (Fig. 7b), bolstering our conclusion—also supported by RTs and occipitoparietal BOLD effects—that subjective reward expectancy increased with elapsed time in the HP condition. Comparing heart-rate timecourses for reward and quit events revealed cardiac deceleration, a well-known correlate of motor preparation[35], prior to quit responses (Fig. 7c). In summary, pre- and post-keypress brain responses (Fig. 6a – f) co-occurred with changes in general somatic arousal, but there was no evidence that arousal effects (as indexed by heart rate) accompanied the trial-onset-locked BOLD effects of theoretical subjective value (Figs. 4 and 5).

Discussion

Decision makers faced with uncertain delay should reappraise awaited rewards as time passes. Depending on the statistics of the environment, the passage of time may either decrease or increase one's estimate of how long a delay remains. This type of dynamic reassessment offers a rationale for sustaining or curtailing persistence.

We elicited either greater or lesser willingness to persist in laboratory environments by manipulating the timing statistics that governed reward delivery. Decision makers calibrated their level of persistence adaptively; this extends previous demonstrations of environment-specific calibration of intertemporal choice behavior[13, 36]. Convergent RT, BOLD and heart-rate data suggested participants encoded the relevant timing statistics, responding more vigorously to more strongly expected rewards[28]. Behavior still fell short of optimality, and an important goal for future work is to determine whether this was due to inexact statistical learning, strong prior expectations, stochastic noise, unmodeled sources of value (e.g., anticipation[37]) or other causes. Future work should also test whether performance would differ if immediate or viscerally tempting rewards were at stake (e.g., appetizing foods instead of money)[1].

The success of the behavioral manipulation enabled us to examine time-dependent brain signals associated with either high or limited behavioral persistence. We observed signals in VMPFC consistent with a dynamic and context-sensitive reassessment of the awaited outcome's subjective value. This effect was identified using both model-guided and exploratory fMRI timecourse analyses, both at the whole-brain level and in ROIs previously implicated in subjective evaluation.

VMPFC and persistence

Persistence toward future rewards has been classically understood to depend on self-regulatory psychological processes that compete with and override more impulsive, reward-sensitive processes[1, 2, 11]. Dual-system psychological models have given rise to the neuroscientific hypothesis that competitive dynamics exist between brain regions subserving cognitive control and reward processing[21–23].

In contrast to this standard view, we have proposed that delay-of-gratification decisions depend on a dynamic reappraisal of the awaited future reward[12, 13]. This account attributes differences in waiting behavior across individuals and situations to factors such as temporal beliefs, perceived outcome values and the perceived cost of time, not merely to differences in the capacity to exert self-control[12]. Here we elicited differences in waiting by manipulating temporal beliefs and obtained evidence for a time-varying representation of subjective value. The hypothesized signal is context dependent, evolving over time in a manner that depends on the timing statistics of the current environment. A corresponding BOLD trajectory was identified in VMPFC, a cortical region regarded as part of a final common pathway in the prospective evaluation of choice alternatives[6]. These results are consistent with the view that persistence depends on the same neural and cognitive processes that guide other forms of reward evaluation and economic choice. This view implies that adaptive persistence depends on accurately representing the value of waiting, and need not depend on the engagement of effortful inhibitory control processes[38]. Our results add to the large body of evidence that VMPFC valuation processes utilize a detailed representation of higher-order task structure[18, 19]. Our findings also extend current conceptions of VMPFC function; while VMPFC activity is known to encode phasic subjective value during discrete choices[3–8], we found that it also tracked subjective value in a temporally extended manner (see Jimura et al.[39] for a related finding).

Our neuroimaging results suggest there is no need to posit antagonistic dynamics between neural reward systems and control systems to explain voluntary persistence (though we cannot, of course, rule out such dynamics in other situations). Our analyses could have detected patterns suggestive of dual-system competition. For example, the analyses in Figs. 4 and 5 and Supplementary Fig. 4 could have detected activity scaling with the difficulty of persistence, but no such effects were found. The analyses in Fig. 6 could have detected a lapse in control-related activity before decisions to quit, but instead the opposite occurred: an ensemble of regions previously implicated in cognitive control—lateral PFC, DMFC, insula, and parietal cortex—increased activity prior to quits, consistent with brain responses found to precede shifts of strategy in other task paradigms[26, 40].

Our findings are more compatible with the hypothesis that cognitive control operates via *value modulation*[9, 17, 20]. The value modulation hypothesis stipulates that control mechanisms in lateral PFC operate by modulating subjective value representations in VMPFC. The hypothesis therefore posits a VMPFC signal that incorporates all relevant information and suffices as a final common pathway to guide decisions, consistent with the present findings. It additionally posits that this signal depends on lateral PFC inputs. On this point our data are mostly silent. We found no evidence for condition-dependent activation trajectories in lateral PFC; nevertheless we assume value computation involves interactions among multiple brain regions, and we cannot exclude the possibility that lateral PFC plays a role.

Value representation during foraging

The problem of calibrating persistence in our willingness-to-wait task is closely analogous to the *patch departure* problem in foraging[24–26]. It has recently been hypothesized that

foraging, which typically involves a succession of accept/reject decisions, imposes fundamentally different information-processing demands from standard multi-alternative economic choice[41]. Recent work has implicated dACC in signaling the value of exiting foraging patches[26] or of shifting away from default alternatives[41], although other findings have questioned this idea[42].

We did not find evidence for continuous, prospective encoding of the value of quitting (analogous to patch departure) in dACC. Such a signal would theoretically have followed an inverted version of the value of waiting (Fig. 3c,d), and could have been detected in either our model-based analysis (as a negative effect) or our exploratory timecourse analyses. We did, however, observe a response in dACC and other frontal and parietal regions in anticipation of quit decisions. This pattern is consistent with general motor preparation as well as with the possibility that decision-related signals in dACC manifest predominantly during overt choice execution[26, 43].

The present results point to a role for VMPFC valuation signals even in a foraging-like situation where decision makers encountered one opportunity at a time and sought to maximize their overall rate of return. VMPFC activity correlated with the value of the current opportunity (waiting for the current token). This finding agrees with the idea that VMPFC encodes a "best minus next-best" comparative value signal[41] even when the "next-best" is the constant background option of moving on to a new opportunity. This parallels previous demonstrations that VMPFC reflects the subjective value of individual options that are evaluated in turn against a fixed reference alternative[7]. Our findings suggest continuity between the valuation mechanisms involved in temporally extended foraging scenarios and multi-option economic choice.

Reward prediction error

The willingness-to-wait task theoretically involves both positive and negative RPE. Positive RPE should accompany reward delivery since, given temporal uncertainty, rewards are not fully predicted at the specific time they are delivered[31]. Conversely, the pre-reward interval (when the reward could have occurred but does not) presumably involves negative RPE[44, 45]. Long delays in the HP environment highlight the dissociability of value and RPE signals. Reward expectancy ramps up over time (Fig. 2c), so nonreward should be associated with progressively larger negative RPE even as the awaited reward's subjective value steadily increases (Fig. 3a). Even though decision makers may be increasingly surprised that the reward did not come *now*, they are also increasingly confident that it will arrive *soon*. One potential explanation for the lack of clear RPE signals in our neuroimaging data might be that, at least at the resolution of fMRI, RPE and subjective value signals were superimposed.

Subjective value is canonically associated with BOLD activity in VMPFC, PCC, and striatum[3–5], and a broad standing question is how these regions might differ in their computational contributions to decision behavior. One possibility is that striatum preferentially encodes RPE[46, 47] whereas VMPFC preferentially encodes prospective decision values[6, 47]. The present findings appear compatible with such a distinction: a dynamic signal of prospective subjective value was observed in VMPFC, but was

significantly less evident in striatum. However, these results will need to be integrated with insights gained using other neuroscientific techniques; recent evidence from direct dopamine recordings suggests striatum may indeed exhibit a ramping pre-reward signal[48], and other work points to an important role for serotonergic neuromodulation in behavioral persistence[49]. It will also be important for future research to assess the fidelity with which VMPFC encodes the individual components of subjective value (Supplementary Fig. 2), to isolate valuation from related factors such as moment-by-moment reward probability[33, 50], and to test the generality of these effects across different magnitudes and types of rewards. Research on these topics will yield an enriched picture of how the brain's valuation mechanisms contend with the complexity of real-world decision environments.

Methods

Participants

The participants were 20 members of the University of Pennsylvania community (age 18–30, mean=22, 11 female). Two additional participants were excluded for head movement (shifts of at least 0.5mm between >5% of adjacent timepoints). Participants were paid a show-up fee ($15/hr) plus rewards earned in the task (median=$19.80). All participants provided informed consent. The procedures were approved by the University of Pennsylvania Internal Review Board. No statistical methods were used to predetermine sample size but our sample size was similar to those reported in previous publications[16, 18, 19, 32, 42, 47].

Task

The task was programmed using Matlab (The MathWorks, Natick, MA) with Psychophysics Toolbox extensions[51, 52]. A circular token, colored green or purple, appeared in the center of the screen, labeled "0¢." After a random delay the token turned blue and its value changed to 30¢. Participants could sell the token anytime by pressing a key with their right hand. The word "SOLD" appeared in red over the token for 1 s. After a 1 s blank screen, a new token appeared. The previous token's value was added to the participant's total earnings, which were displayed only at the end of each scanning run. Setting the token's initial value to 0¢ meant that, unlike earlier work using this paradigm[13], participants received no immediate reward upon quitting. This served to simplify the task without significantly altering either its incentive structure or the resulting pattern of behavior.

A white progress bar marked the amount of time the current token had been on the screen. The bar's full length corresponded to 100 s. It grew continuously from the left and reset when a new token appeared. The progress bar was included to reduce interval-timing demands and discourage a strategy of covertly counting time.

The scanning session was divided into four 10-min runs. New tokens were presented until time was up. Each run presented one timing environment (i.e., token color). The two environments alternated in successive runs. The order of environments and the mapping of token color to environment were counterbalanced across participants.

Each participant completed a preliminary behavioral training session consisting of 4 10-min runs alternating between the HP and LP environments. Participants were explicitly instructed that the green and purple tokens might differ in their timing, but that they had to learn the nature of the differences from direct experience and were free to adopt any behavioral strategy they preferred. During behavioral training (but not during scanning) the screen displayed the time left in the 10-min run and the amount earned so far, to help ensure that participants understood the structure of the task. Each token during behavioral training was worth 10¢. Participants explored the task environments during training, waiting through full 90s delays in the LP condition on a median of 3.5 trials (IQR 1 to 5.5; >0 for 18/20 subjects). Participants completed two additional 5-min runs (one per condition) outside the scanner just before the fMRI session. Waiting behavior over time across training, practice, and fMRI sessions is plotted in Supplementary Fig. 1.

Participants would have faced fundamentally the same trial-by-trial decision problem if they had received explicit information about the probabilistic contingencies in lieu of experience-based training (cf. Luhmann et al.[53]). However, there is evidence that probabilistic information is encoded differently when learned from description vs. direct experience[54–56]; our training procedure was designed to involve the type of experience-based, implicit statistical learning that is thought to guide beliefs and expectations in real-world domains[29], including ecological foraging environments. Future work might introduce explicit information to help assess whether deviations from optimal behavior were due to inexact encoding of the relevant probabilities.

The delay duration on each trial was randomly drawn from a discrete probability distribution (Fig 1b). In the HP environment delays were drawn uniformly from the values 5, 10, 15, 20, 25, 30, 35, and 40s. In the LP environment delays were set to 90s with probability 0.5, and otherwise drawn uniformly from the values 5, 10, 15, and 20s. By design, reward probabilities were identical between the two environments for the first 20s of the delay, the period of greatest interest in our neuroimaging analyses. We imposed longer delays here than in our previous work[13] in order to obtain fluctuations in subjective value across a time period on the order of 30s, which is well suited for detecting BOLD effects (this corresponds to the time scale of a blocked design with ~15 s blocks; see further discussion and simulation results below).

Delays were sampled in a pseudorandom manner that approximately balanced the first-order transition statistics between delays in successive trials. This helped ensure that the scheduled delays were representative of the ground-truth distribution, while avoiding the negative autocorrelation that would result from strictly balanced frequencies.

The HP environment was richer by design, with all participants receiving more rewards in the HP environment (median=44, IQR 41 to 46.5) than in the LP environment (median=25, IQR 21 to 26). The difference in overall richness was not the factor that determined the ideal behavioral strategy (one could design richer LP environments and poorer HP environments[13]), but emerged here as a side-effect of our decision to match the sizes of individual rewards and the reward probabilities over the first 20s. These design choices

maximized the comparability of the two conditions for purposes of our neuroimaging contrasts.

We quantified behavioral persistence using Kaplan-Meier survival curves[57], which estimated the probability of "surviving" various lengths of time without quitting. This technique accommodated the fact that reward delivery censored observed waiting times[13].

Modeling ideal performance

The rate-maximizing strategy was to wait through all delays in the HP environment (up to 40 s), but to give up after 20 s in the LP environment. We determined this by calculating the expected rate of return for various giving-up times (Fig. 1c). This calculation follows previous work[13], and has precedent in stochastic foraging models[25].

The reward's arrival time t_{reward} is a random variable. For a policy of quitting at time T, let p_T equal the probability of receiving the reward, $p_T = Pr(t_{reward} \quad T)$, and let τ_T equal the expected delay if the reward is received, $\tau_T = E(t_{reward} \mid t_{reward} \quad T)$. Each trial's expected rate of return, in ¢/s, is:

$$R_T = \frac{30 p_T}{\tau_T p_T + T(1 - p_T) + 2} \quad (1)$$

The numerator is a trial's expected gain in cents and the denominator is a trial's expected cost in seconds, assuming a 30¢ reward and a 2 s inter-trial interval. The goal is to find the value of T that maximizes R_T. We use R^* to denote the best available rate of return. Fig. 1c plots R_T as a function of T. The best policy in the HP environment was to wait 40s ($R^* = 1.22$¢/s), whereas the best policy in the LP environment was to wait 20s ($R^* = 0.82$¢/s).

Modeling subjective value as a function of elapsed time

At each point in a trial, the token's subjective value depended on three factors: (1) the expected earnings from that token, (2) the expected additional time to be spent on that token, and (3) the monetary value of time, which corresponds to R^* from above. We denote the expected earnings as $a_T(t)$ and the expected time as $b_T(t)$. Each of these depends jointly on the current elapsed time t and the intended future quitting time T. For given values of t and T, the expected return is:

$$g_T(t) = a_T(t) - (R^*) b_T(t) \quad (2)$$

The current trial's subjective value (denoted "potential" in the model's original formulation[25]) equals the maximum value of g_T across all possible quitting times:

$$g(t) = \max_{T} g_T(t) \quad (3)$$

Put differently, $g(t)$ is the expected net return in the remainder of the current trial, accounting for the cost of time, under the best available waiting policy. Its minimum is zero

because there is always an option to quit immediately (we treat the ITI as part of the subsequent trial). The decision maker should continue waiting if $g(t) > 0$.

Fig. 3a shows $g(t)$ as a function of t in each environment (see Supplementary Fig. 2 for decomposition of g(t) into its components). The function approaches 30¢ at the last possible reward time, when a 30¢ reward is expected with no further delay. The best strategy is to wait up to 40 s in the HP environment but quit at 20 s in the LP environment. If a decision maker in the LP environment were to have waited 53.5 s already it would be better at that point to continue waiting for the reward that was sure to arrive at 90 s. We obtained very similar results if we used each participant's actual environment-specific reward rate in place of the theoretical maximum, $R*$ (Supplementary Fig. 6).

Behavior could be well characterized as a stochastic function of theoretical subjective value. To evaluate this we represented each subject's behavior as a series of pseudo-choices between waiting and quitting, placed every 1s throughout all delay intervals in the experiment. We then modeled pseudo-choice outcomes (1=wait, 0=quit) as a function of subjective value and a constant intercept in subject-wise logistic regressions. Subject-wise maximum-likelihood coefficients were tested at the group level using a Wilcoxon signed-rank test. We additionally used likelihood-ratio tests at the single-subject level to compare the full model to the (nested) intercept-only model, and tested the resulting z statistics at the group level. Finally, we tested an alternative model that, in place of subjective value, coded the HP and LP conditions categorically. This model had the same number of parameters as the subjective value model and could represent the possibility that participants merely quit more often in the LP than HP environment. Subject-wise differences in model deviance were tested against zero using a group-level Wilcoxon signed-ranks test.

Allowing for endogenous temporal uncertainty did not substantially alter the theoretical results described above. Fig. 2c displays hypothetical continuous hazard functions allowing for subjective uncertainty in time-interval perception[58]. For an interval of true duration t, subjective uncertainty is typically well characterized by a Gaussian distribution with mean μ $= t$ and standard deviation σ $= t \times CV$, where CV is a fixed *coefficient of variation*. We modeled temporal uncertainty by converting each discrete distribution in Fig. 1b to a Gaussian mixture distribution. A Gaussian component was placed at each possible reward time t, with μ $= t$, σ $= t \times CV$, and weight equal to $Pr(t_{reward}=t)$. We set CV=0.16 on the basis of human behavioral findings (the median CV from Table 2 of Rakitin et al.[59] after converting the unit of variability from full-width-at-half-maximum to SD). The continuous functions in Fig. 2c are scaled by a factor of 5 for comparability with the corresponding discrete functions.

Blurring the ground-truth timing distributions to allow for subjective uncertainty did not change any of our model-based theoretical predictions. If rates of return (Fig. 1c) were calculated using the Gaussian mixture distribution, the best policy was to wait 40 s in the HP environment and 22.1 s in the LP environment. Endogenous uncertainty smoothed the theoretical subjective value functions (Fig. 3a) without altering their general shape.

MRI data acquisition and preprocessing

MRI data were acquired on a 3T Siemens Trio with a 32-channel head coil. Functional data were acquired using a gradient-echo echoplanar imaging (EPI) sequence (3mm isotropic voxels, 64×64 matrix, 44 axial slices tilted 30° from the AC-PC plane, TR=2500 ms, TE=25 ms, flip angle=75°). There were 4 runs, each with 246 images (10 min, 15 s). At the end of the session we acquired matched fieldmap images (TR=1000 ms, TE=2.69 and 5.27 ms, flip angle=60°) and a T1-weighted MPRAGE structural image (0.9375×0.9375×1 mm voxels, 192×256 matrix, 160 axial slices, TI=1100 ms, TR=1630 ms, TE=3.11 ms, flip angle=15°).

Data were preprocessed using FSL[60–63] and AFNI[64, 65] software. Functional data were temporally aligned to midpoint of each acquisition (AFNI's *3dTshift*), motion corrected (FSL's *MCFLIRT*), undistorted and warped to MNI space (see below), outlier-attenuated (AFNI's *3dDespike*), smoothed with a 6 mm FWHM Gaussian kernel (FSL's *fslmaths*), and intensity-scaled by a single grand-mean value per run. To warp the data to MNI space, functional data were aligned to the structural image (FSL's *FLIRT*) using boundary-based registration[66], simultaneously incorporating fieldmap-based geometric undistortion. Separately, the structural image was nonlinearly coregistered to the MNI template (FSL's *FLIRT* and *FNIRT*). The two transformations were concatenated and applied to the functional data.

fMRI analysis

Voxelwise general linear models (GLMs) were fit using ordinary least squares (AFNI's *3dDeconvolve*). GLMs were estimated for each subject individually using data concatenated across the 4 runs. There were 12 baseline terms per run: a constant, 5 low-frequency drift terms (first-through-fifth-order Legendre polynomials), and 6 motion parameters.

Event-related BOLD signal timecourses were flexibly estimated by fitting piecewise linear splines ("tent" basis functions). For trial-onset-locked timecourses, basis functions were centered every 2.5 s beginning at 2.5 s and ending 1 s before the end of each trial (for example, the basis function regressor corresponding to "10 s" had a peak 10 s after trial onset for every trial that lasted at least 11 s). For reward-related and quit-related timecourses, basis functions were centered every 2.5 s from 12.5 s before to 12.5 s after the event.

We conducted simulations to confirm the validity of our analysis procedures. We calculated theoretical subjective value over the course of each subject's entire experimental session using the actual timing of task events together with the ideal model in Fig. 3a. These full-session timecourses were convolved with a hemodynamic response function (HRF) to generate subject-specific synthetic BOLD timecourses representing idealized theoretical predictions. In order to verify that the theoretical signal had a suitable time scale and could be distinguished from baseline drift, we fit these synthetic BOLD timecourses in a GLM that contained only the constant and drift terms. For each subject, the residuals were highly correlated with a merely de-meaned version of the original synthetic BOLD timecourses (median r^2=0.90, IQR 0.88 to 0.93), indicating that the theoretical signal could indeed be clearly distinguished from baseline fluctuations. Next we used the synthetic BOLD

timecourses as inputs to the analysis procedure described above for estimating trial-onset-locked timecourses. The resulting timecourses, shown in averaged form in Fig. 3d, constituted our subject-by-subject theoretical predictions.

The model-based analysis was performed voxelwise on all 20 subjects across 2.5–30 s from trial onset. Each subject's empirically estimated difference timecourse (HP minus LP) was regressed on the theoretical difference timecourse (Fig. 3d) together with a constant intercept. Using the simplified HRF-convolved theoretical timecourses in Fig. 3c yielded equivalent results. Timepoints lacking data in either environment for a given subject were omitted (this resulted in the omission of 3 timepoints for one subject; see Supplementary Fig. 3). We adopted a two-step approach (first estimating the timecourses and then submitting them to the model-based contrast) so that included timepoints were weighted uniformly. Otherwise, early timepoints, which were sampled more frequently (Supplementary Fig. 3), would have received greater weight, and the pattern of timepoint weighting could have differed between environments for individual subjects. Contrast coefficients were tested against zero at the group level in 2-tailed voxelwise t-tests.

An additional open-ended analysis tested for condition-by-timepoint interactions in the trial-onset-locked BOLD timecourses (using n=19 participants with complete data; Supplementary Fig. 3). The main effect of timepoint is of limited interest because it captures nonspecific effects of time-from-keypress; similarly, the main effect of condition is uninformative because the two conditions were presented in separate runs with independent baselines. The condition-by-timepoint interaction tests for a difference in BOLD trajectories between the two environments without constraining the form of the difference. In a repeated-measures framework this is equivalent to testing the main effect of timepoint on the *difference* in signal between the two environments. Accordingly, we performed a voxelwise one-way repeated-measures ANOVA on the difference timecourses (HP minus LP) at the group level. An equivalent procedure was used to compare BOLD timecourses aligned to reward-related and quit-related keypresses.

The RPE analysis was limited to the HP environment, in which the sustained rise in reward expectancy supported clear predictions. Within a GLM we estimated FIR coefficients for the peri-reward timecourse (from 7.5 s before to 10 s after each reward). Eight terms modeled the mean timecourse, and another 8 terms modeled amplitude modulation at each timepoint as a function of the preceding delay duration. We then computed a contrast of the modulatory effect for 3 post-reward timepoints (2.5 to 7.5 s) minus 3 earlier timepoints (–5 to 0 s). The value of this contrast reflected modulation of the phasic reward response as a function of preceding delay time, over and above any nonspecific effect of elapsed time on the pre-reward baseline.

All whole-brain, group-level analyses assessed statistical significance on the basis of cluster mass, with the cluster-defining threshold set to the nominal $p<0.01$ level. Corrected p-values were determined using permutation testing[67] (FSL's *randomise*; 5000 iterations), and results were thresholded at corrected $p<0.05$. For F tests, each random iteration shuffled timepoints within subject. For one-sample t-tests, each iteration randomly sign-flipped individual subjects' coefficient maps.

Heart rate data acquisition and analysis

Pulse oximetry data were recorded at 50 Hz using the MRI system's built-in oximeter, which also performed automatic heartbeat detection. Timestamped data were successfully recorded for 17 of the 20 participants. Heartbeat times were converted to inter-beat interval (IBI). IBI values farther than 30% above or below the grand median were treated as missing (median = 1.6% of points removed; IQR 0.5 to 3.6%). Since IBI varied across individuals (median=820 ms; IQR 760 to 950 ms), IBI values were converted to a percentage of the individual's grand median. Mean perievent timecourses were calculated on a 0.25 s grid for each subject and event type. For comparisons, timecourses for two event types were subtracted to yield single-subject difference timecourses, which were then tested at the group level for significant excursions from zero. Entire timecourses were tested using cluster-based control for multiple comparisons. Cluster size was defined as the number of adjacent timepoints with nominal $p<0.05$ in single-timepoint Wilcoxon signed-rank tests. A cluster was assigned a corrected p-value based on its percentile in the empirical null distribution for cluster size, which was obtained via permutation testing (10,000 iterations with randomized sign-flipping of individual subjects' difference timecourses).

A supplementary methods checklist is available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Mischel W, Ebbesen EB. Attention in delay of gratification. Journal of Personality and Social Psychology. 1970; 16:329–337.

2. Baumeister RF, Vohs KD, Tice DM. The strength model of self-control. Current Directions in Psychological Science. 2007; 16:351–355.

3. Bartra O, McGuire JT, Kable JW. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage. 2013; 76:412–427. [PubMed: 23507394]

4. Clithero JA, Rangel A. Informatic parcellation of the network involved in the computation of subjective value. Social Cognitive and Affective Neuroscience. 2013

5. Liu X, Hairston J, Schrier M, Fan J. Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. Neuroscience and Biobehavioral Reviews. 2011; 35:1219–1236. [PubMed: 21185861]

6. Levy DJ, Glimcher PW. The root of all value: A neural common currency for choice. Current Opinion in Neurobiology. 2012; 22:1027–1038. [PubMed: 22766486]

7. Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. Nature Neuroscience. 2007; 10:1625–1633. [PubMed: 17982449]

8. Kable JW, Glimcher PW. An "as soon as possible" effect in human intertemporal decision making: Behavioral evidence and neural mechanisms. Journal of Neurophysiology. 2010; 103:2513–2531. [PubMed: 20181737]

9. Hare TA, Camerer CF, Rangel A. Self-control in decision-making involves modulation of the vmPFC valuation system. Science. 2009; 324:646–648. [PubMed: 19407204]

10. Mischel, W.; Ayduk, O.; Mendoza-Denton, R. Sustaining delay of gratification over time: A hot-cool systems perspective. In: Loewenstein, G.; Read, D.; Baumeister, RF., editors. Time and decision: Economic and psychological perspectives on intertemporal choice. New York: Russell Sage Foundation; 2003. p. 175-200.

11. Metcalfe J, Mischel W. A hot/cool-system analysis of delay of gratification: dynamics of willpower. Psychological Review. 1999; 106:3–19. [PubMed: 10197361]

12. McGuire JT, Kable JW. Rational temporal predictions can underlie apparent failures to delay gratification. Psychological Review. 2013; 120:395–410. [PubMed: 23458085]

13. McGuire JT, Kable JW. Decision makers calibrate behavioral persistence on the basis of time-interval experience. Cognition. 2012; 124:216–226. [PubMed: 22533999]

14. Rachlin, H. The science of self-control. Cambridge, MA: Harvard University Press; 2000.

15. Dasgupta P, Maskin E. Uncertainty and hyperbolic discounting. American Economic Review. 2005; 95:1290–1299.

16. Kim H, Shimojo S, O'Doherty JP. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. Cerebral Cortex. 2011; 21:769–776. [PubMed: 20732900]

17. Hare TA, Malmaud J, Rangel A. Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. Journal of Neuroscience. 2011; 31:11077–11087. [PubMed: 21795556]

18. Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. Journal of Neuroscience. 2006; 26:8360–8367. [PubMed: 16899731]

19. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron. 2011; 69:1204–1215. [PubMed: 21435563]

20. Hutcherson CA, Plassmann H, Gross JJ, Rangel A. Cognitive regulation during decision making shifts behavioral control between ventromedial and dorsolateral prefrontal value systems. Journal of Neuroscience. 2012; 32:13543–13554. [PubMed: 23015444]

21. Casey BJ, et al. Behavioral and neural correlates of delay of gratification 40 years later. Proceedings of the National Academy of Sciences. 2011; 36:14998–15003.

22. Figner B, et al. Lateral prefrontal cortex and self-control in intertemporal choice. Nature Neuroscience. 2010; 13:538–539. [PubMed: 20348919]

23. Heatherton TF, Wagner DD. Cognitive neuroscience of self-regulation failure. Trends in Cognitive Sciences. 2011; 15:132–139. [PubMed: 21273114]

24. Charnov EL. Optimal foraging, the marginal value theorem. Theoretical Population Biology. 1976; 9:129–136. [PubMed: 1273796]

25. McNamara J. Optimal patch use in a stochastic environment. Theoretical Population Biology. 1982; 21:269–288.

26. Hayden BY, Pearson JM, Platt ML. Neuronal basis of sequential foraging decisions in a patchy environment. Nature Neuroscience. 2011; 14:933–939. [PubMed: 21642973]

27. Fawcett TW, McNamara JM, Houston AI. When is it adaptive to be patient? A general framework for evaluating delayed rewards. Behavioural Processes. 2012; 89:128–136. [PubMed: 21920413]

28. Nickerson RS. Response time to the second of two successive signals as a function of absolute and relative duration of intersignal interval. Perceptual and Motor Skills. 1965; 21:3–10. [PubMed: 5828404]

29. Griffiths TL, Tenenbaum JB. Optimal predictions in everyday cognition. Psychological Science. 2006; 17:767–773. [PubMed: 16984293]

30. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. Journal of Neuroscience. 1996; 16:1936–1947. [PubMed: 8774460]

31. Fiorillo CD, Newsome WT, Schultz W. The temporal precision of reward prediction in dopamine neurons. Nature Neuroscience. 2008; 11:966–973. [PubMed: 18660807]

32. Cui X, Stetson C, Montague PR, Eagleman DM. Ready…go: Amplitude of the FMRI signal encodes expectation of cue arrival time. PLoS biology. 2009; 7:e1000167. [PubMed: 19652698]

33. Bueti D, Bahrami B, Walsh V, Rees G. Encoding of temporal probabilities in the human brain. Journal of Neuroscience. 2010; 30:4343–4352. [PubMed: 20335470]

34. Tootell RBH, et al. The retinotopy of visual spatial attention. Neuron. 1998; 21:1409–1422. [PubMed: 9883733]

35. Lacey, JI.; Lacey, BC. Some autonomic-central nervous system interrelationships. In: Black, P., editor. Physiological correlates of emotion. New York: Academic Press; 1970. p. 205-227.

36. Schweighofer N, et al. Humans can adopt optimal discounting strategy under real-time constraints. PLoS Computational Biology. 2006; 2:e152. [PubMed: 17096592]

37. Loewenstein G. Anticipation and the valuation of delayed consumption. The Economic Journal. 1987; 97:666–684.

38. Duckworth AL, Gendler TS, Gross JJ. Self-control in school-age children. Educational Psychologist. 2014; 49:199–217.

39. Jimura K, Chushak MS, Braver TS. Impulsivity and self-control during intertemporal decision making linked to the neural dynamics of reward value representation. Journal of Neuroscience. 2013; 33:344–357. [PubMed: 23283347]

40. Helfinstein SM, et al. Predicting risky choices from brain activity patterns. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:2470–2475. [PubMed: 24550270]

41. Rushworth MFS, Kolling N, Sallet J, Mars RB. Valuation and decision-making in frontal cortex: One or many serial or parallel systems? Current Opinion in Neurobiology. 2012; 22:946–955. [PubMed: 22572389]

42. Shenhav A, Straccia MA, Cohen JD, Botvinick MM. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. Nature Neuroscience. 2014; 17:1249–1254. [PubMed: 25064851]

43. Blanchard TC, Hayden BY. Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014; 34:646–655. [PubMed: 24403162]

44. McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron. 2003; 38:339–346. [PubMed: 12718866]

45. Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience. 1998; 1:304–309. [PubMed: 10195164]

46. Berns GS, McClure SM, Pagnoni G, Montague PR. Predictability modulates human brain response to reward. Journal of Neuroscience. 2001; 21:2793–2798. [PubMed: 11306631]

47. Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. Journal of Neuroscience. 2008; 28:5623–5630. [PubMed: 18509023]

48. Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature. 2013; 500:575–579. [PubMed: 23913271]

49. Miyazaki KW, et al. Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. Current Biology. 2014; 24:2033–2040. [PubMed: 25155504]

50. Janssen P, Shadlen MN. A representation of the hazard rate of elapsed time in macaque area LIP. Nature Neuroscience. 2005; 8:234–241. [PubMed: 15657597]

51. Brainard DH. The psychophysics toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

52. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision. 1997; 10:437–442. [PubMed: 9176953]

53. Luhmann CC, Chun MM, Yi D-J, Lee D, Wang X-J. Neural dissociation of delay and uncertainty in intertemporal choice. Journal of Neuroscience. 2008; 28:14459–14466. [PubMed: 19118180]

54. Ungemach C, Chater N, Stewart N. Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? Psychological Science. 2009; 20:473–479. [PubMed: 19399978]

55. FitzGerald THB, Seymour B, Bach DR, Dolan RJ. Differentiable neural substrates for learned and described value and risk. Current Biology. 2010; 20:1823–1829. [PubMed: 20888231]

56. Hertwig R, Barron G, Weber EU, Erev I. Decisions from experience and the effect of rare events in risky choice. Psychological Science. 2004; 15:534–539. [PubMed: 15270998]

57. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958; 53:457–481.

58. Gibbon J. Scalar expectancy theory and Weber's law in animal timing. Psychological Review. 1977; 84:279–325.

59. Rakitin BC, et al. Scalar expectancy theory and peak-interval timing in humans. Journal of Experimental Psychology: Animal Behavior Processes. 1998; 24:15–33. [PubMed: 9438963]

60. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Medical Image Analysis. 2001; 5:143–156. [PubMed: 11516708]

61. Smith SM, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004; 23:S208–S219. [PubMed: 15501092]

62. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage. 2012; 62:782–790. [PubMed: 21979382]

63. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage. 2002; 17:825–841. [PubMed: 12377157]

64. Cox RW. AFNI: What a long strange trip it's been. NeuroImage. 2012; 62:743–747. [PubMed: 21889996]

65. Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29:162–173. [PubMed: 8812068]

66. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. NeuroImage. 2009; 48:63–72. [PubMed: 19573611]

67. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping. 2002; 15:1–25. [PubMed: 11747097]
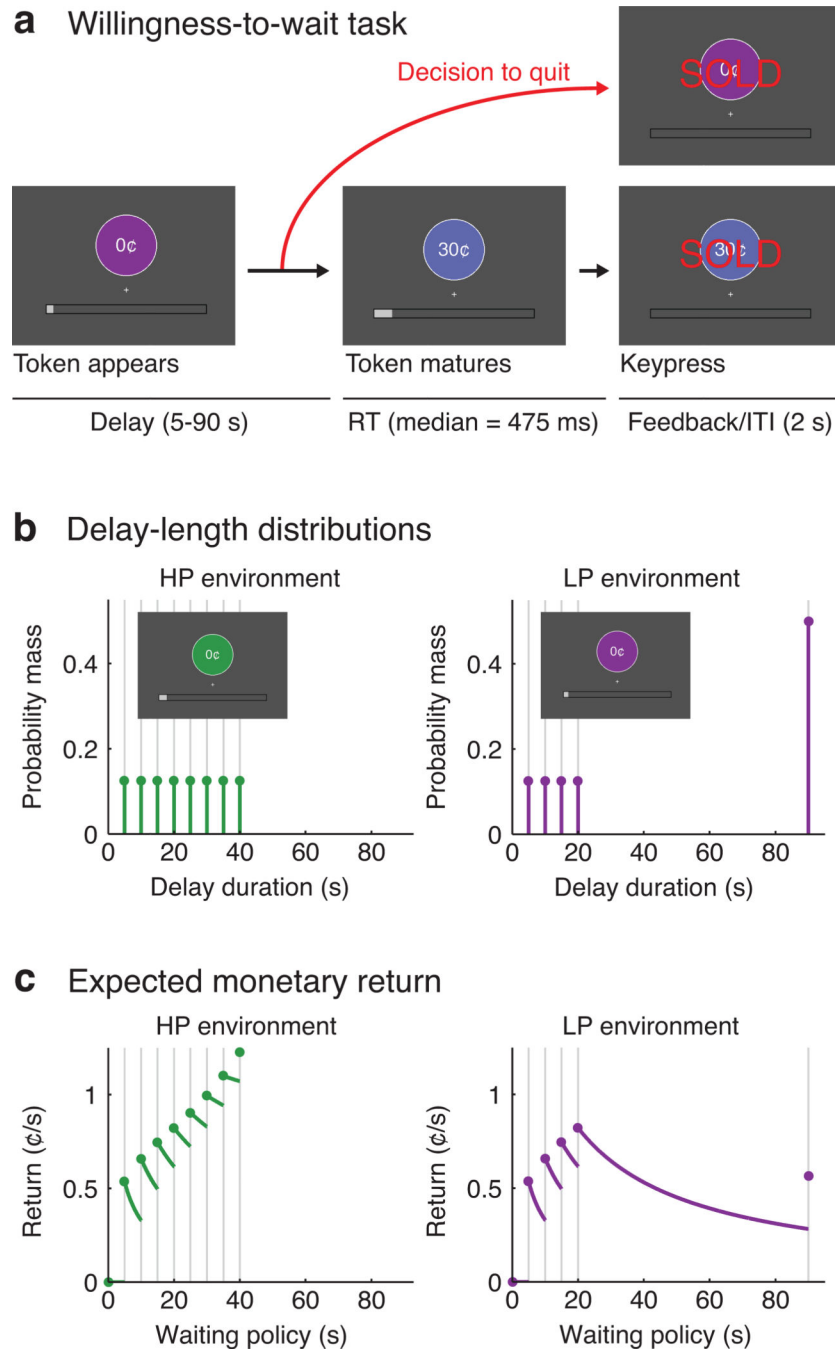
Figure 1.
Experimental task and timing conditions. A: Schematic of the willingness-to-wait task. B: Discrete probability distributions governing the scheduled delay times in each environment. C: Expected monetary rates of return under various waiting policies, where each policy is defined by a giving-up time. The reward-maximizing policy was to wait up to 40s in the HP environment (i.e., never to quit), but only up to 20s in the LP environment. These rates of return are contingent on the fixed 2s inter-trial interval (ITI).

a

b

c

d
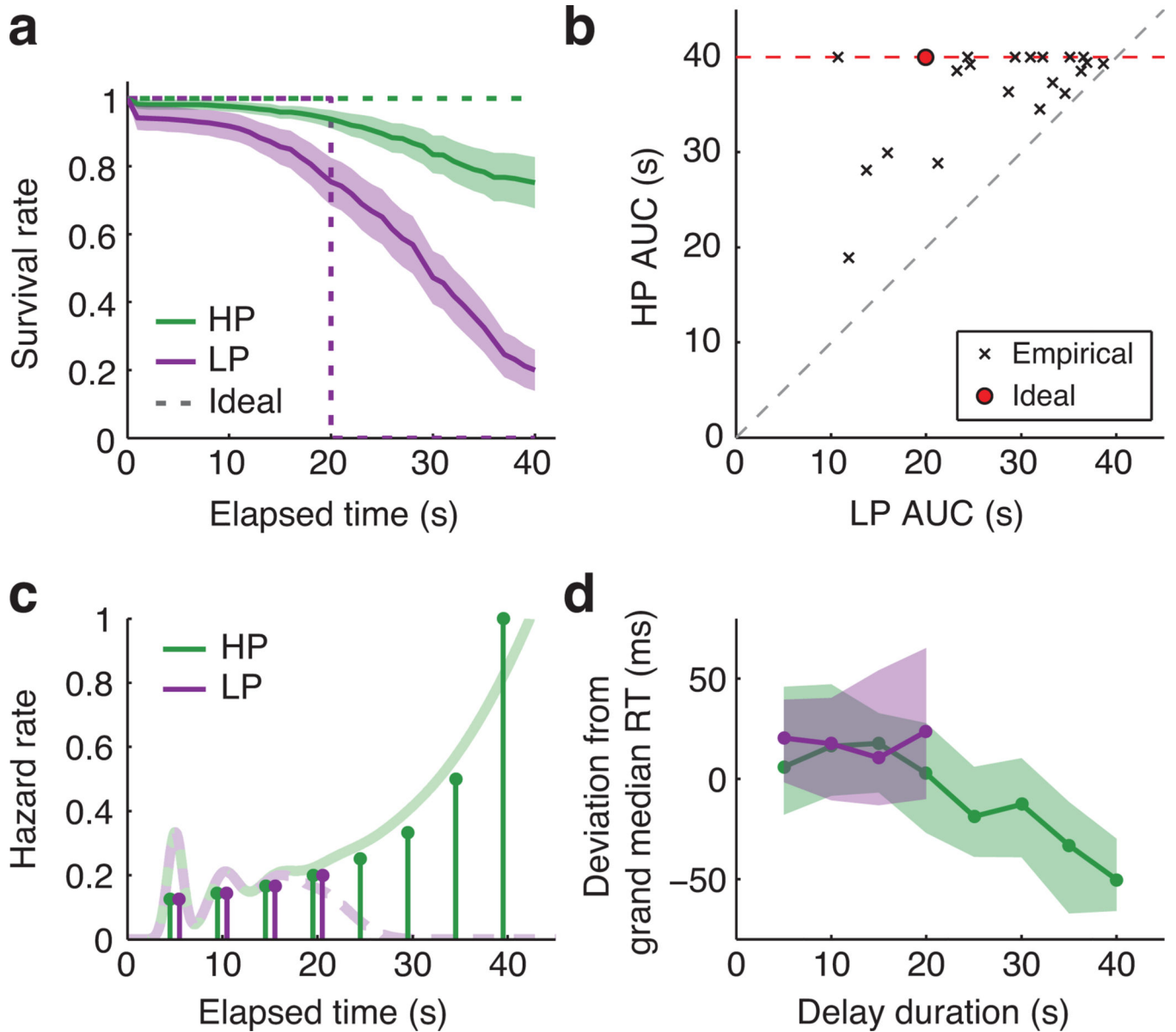
Figure 2.

Behavioral results. A: Survival curves reflecting the probability that a participant was still waiting at each elapsed time, provided that the reward had not yet been delivered. Empirical survival curves were averaged across subjects at 1 s intervals (+/− SEM). Ideal performance is plotted for reference (dashed lines). B: Area under the curve (AUC) values calculated from individual participants' survival curves. The maximum possible value was 40s. Red point marks ideal performance. All 20 participants persisted more in the HP environment. C: Stem plots show the ground-truth hazard rate for reward in each environment: i.e., the probability of the reward arriving at each time, conditional on not having arrived already. Faded lines illustrate hypothetical continuous hazard functions incorporating endogenous temporal uncertainty (see *Methods*). D: Reward RT at each delay (median and IQR of subject-wise medians). RTs are expressed as deviations from each subject's grand-median
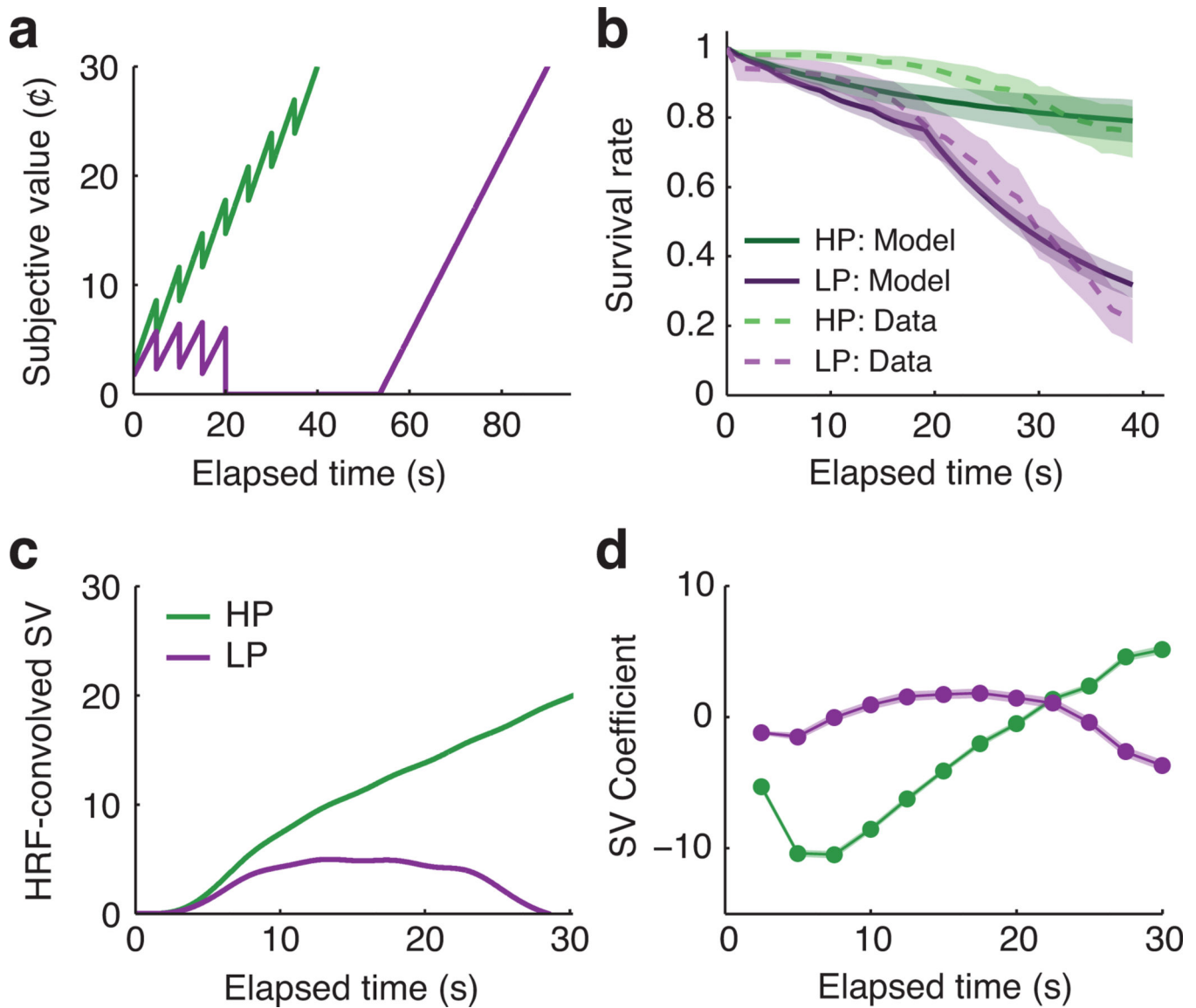
RT (median=475ms, IQR 450 to 506ms) to display within-subject effects. RTs for 5–20s delays did not differ between the environments (HP median=472ms, IQR 454 to 538; LP median=494ms, IQR 443 to 522).
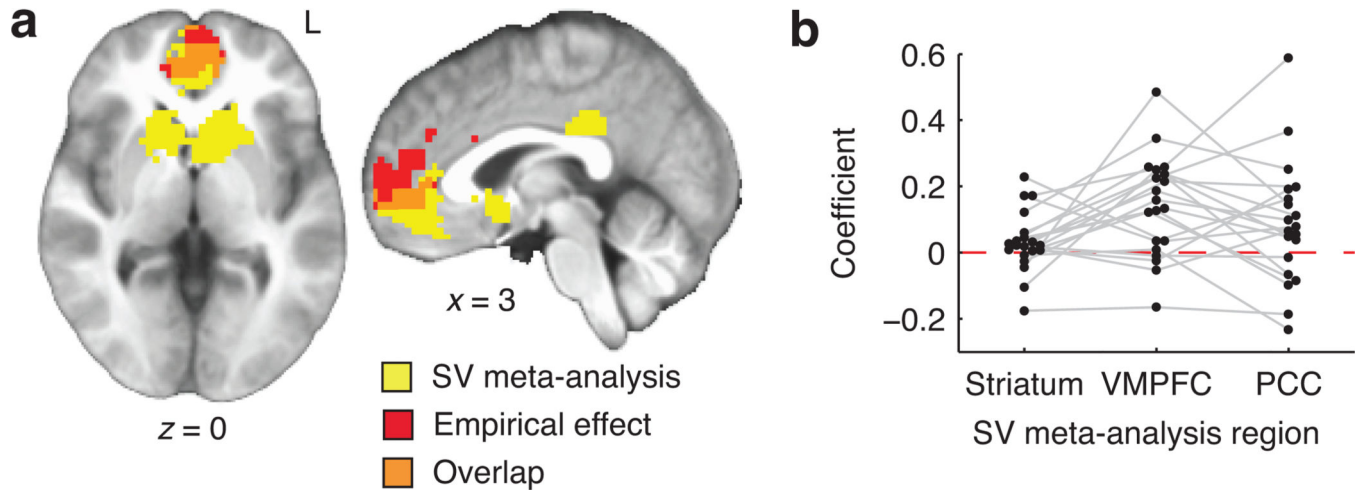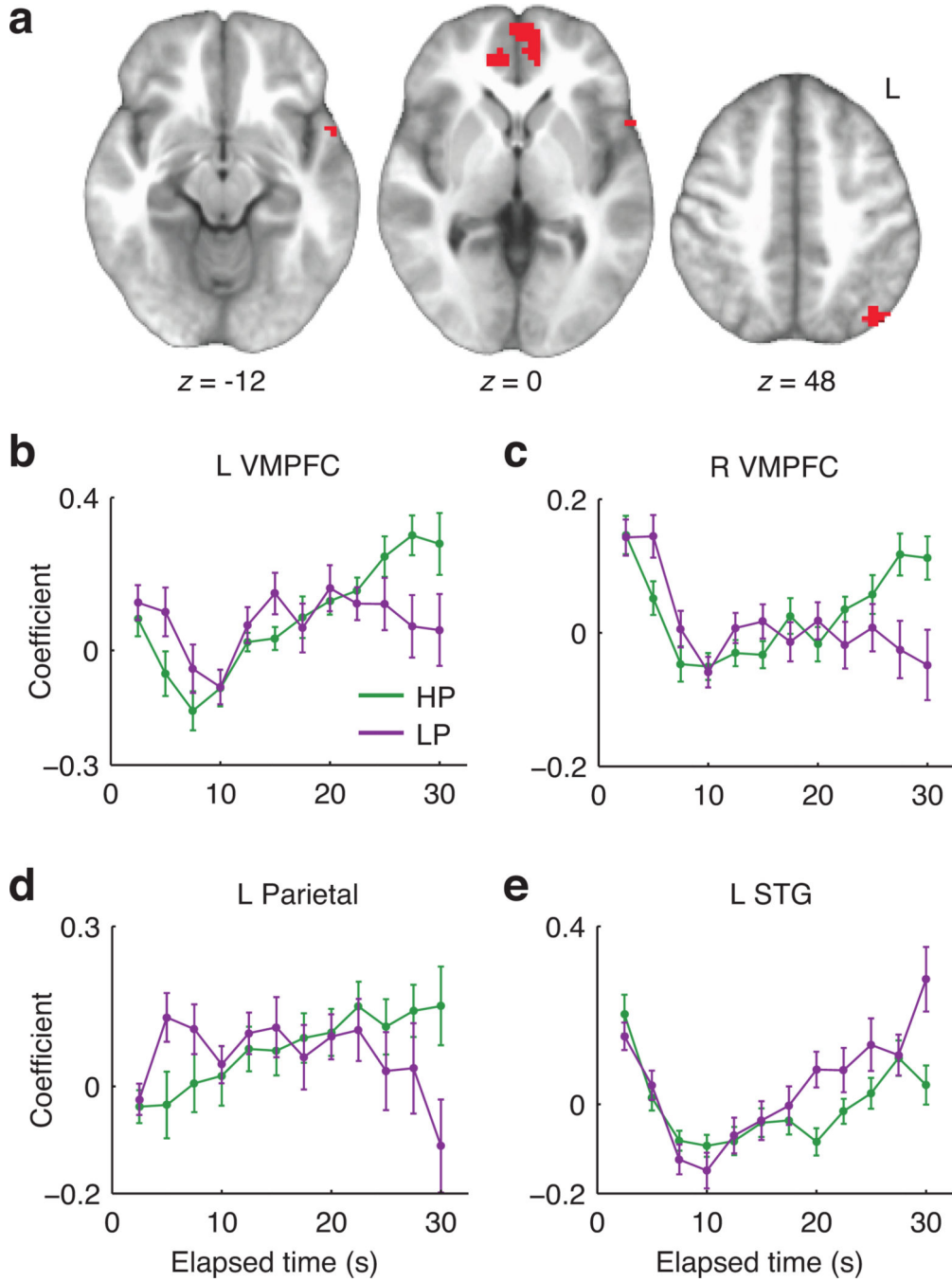
Figure 3.
Theoretical subjective value of the awaited token as a function of elapsed time in each environment. A: A token's subjective value increased over time in the HP environment but not in the LP environment. These timecourses are based on the discrete ground-truth timing distributions and would be smoothed by subjective temporal uncertainty. B: Simulated behavior from a model in which subjective value linearly influenced the log-odds of continuing to wait (mean +/− SEM of subject-wise model fits). Data from Fig. 2a are overlaid for reference. C: Subjective value timecourses convolved with a canonical hemodynamic response function (HRF). D: Predicted BOLD timecourses obtained by applying our fMRI analysis to idealized synthetic data (mean +/− SEM of individual subject results). Visual differences from Panel C reflect that (1) the HP and LP environments had independent baselines, and (2) there was a small degree of carryover across trials. In spite of these differences the theoretical difference timecourses (HP minus LP) were highly correlated between Panels C and D (median r^2=0.88, IQR 0.84 to 0.89).

Figure 4.
Model-based contrast results. A: Whole-brain analysis. Displayed in red is the VMPFC cluster that showed a significant relationship with the theoretical subjective value timecourses in Fig. 3d. In yellow, for reference, are regions identified in a previous meta-analysis of valuation effects (the regions reported in Fig. 3D of Bartra et al.[3]). Overlap was observed in VMPFC, though not in PCC or striatum. B: Model-based contrast values for each participant, spatially averaged within meta-analytic ROIs. Subjective value effects were significantly positive in VMPFC, and significantly greater in VMPFC than striatum.

Figure 5.
Model-free analysis of trial-onset-locked BOLD timecourses. A: Clusters showing a significant timepoint-by-environment interaction (Table 1b). B–E: Spatially averaged signal timecourses for significant clusters (mean +/− SEM), illustrating the form of the observed interactions. Although voxel selection effects would distort follow-up inferential tests of these timecourses, we descriptively summarized their resemblance to our theoretical predictions in terms of the correlation between the average theoretical (Fig. 3d) and

observed HP-minus-LP difference timecourses. The resulting Pearson *r* values were 0.91, 0.89, 0.90, and −0.68 for the results in Panels B–E, respectively.
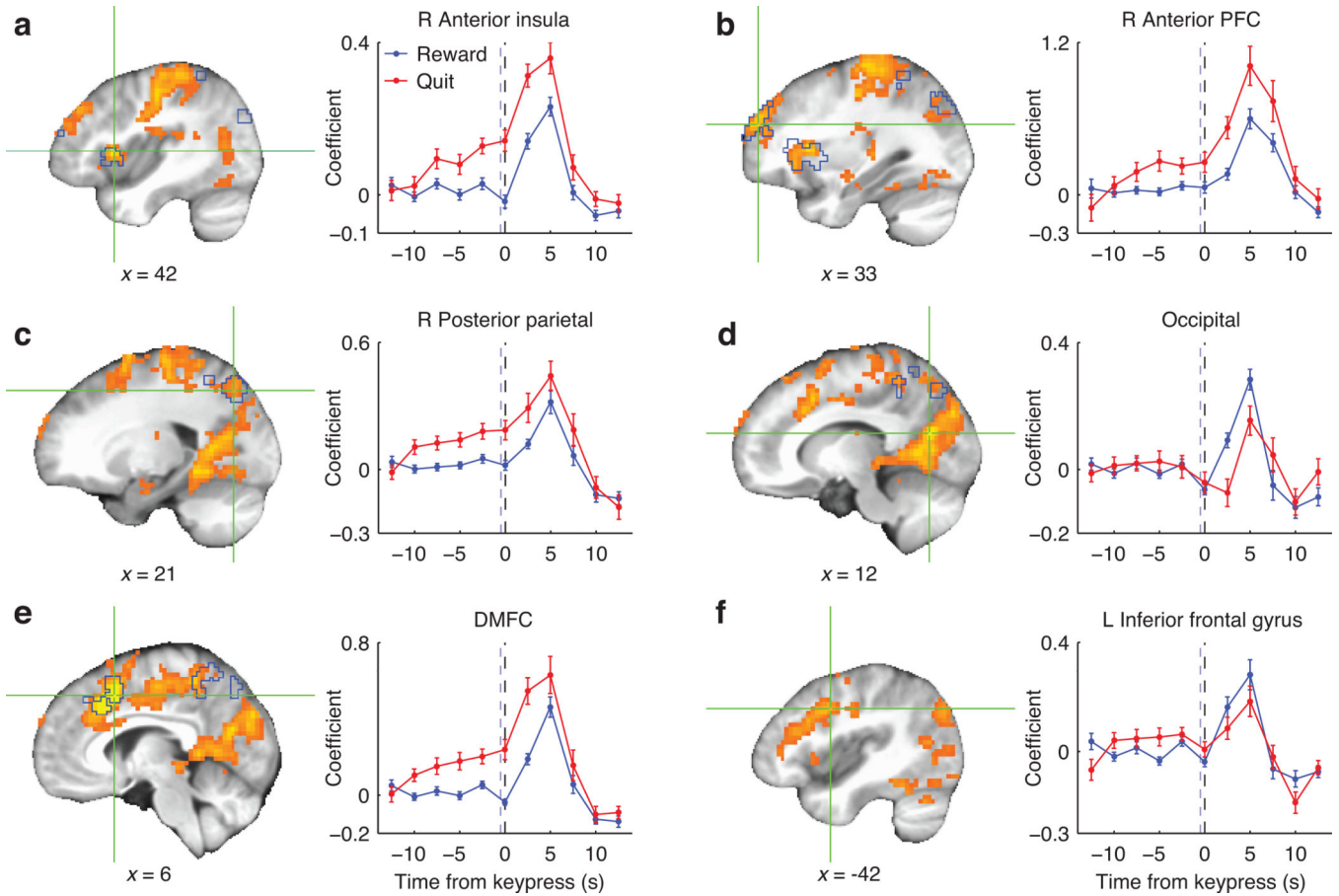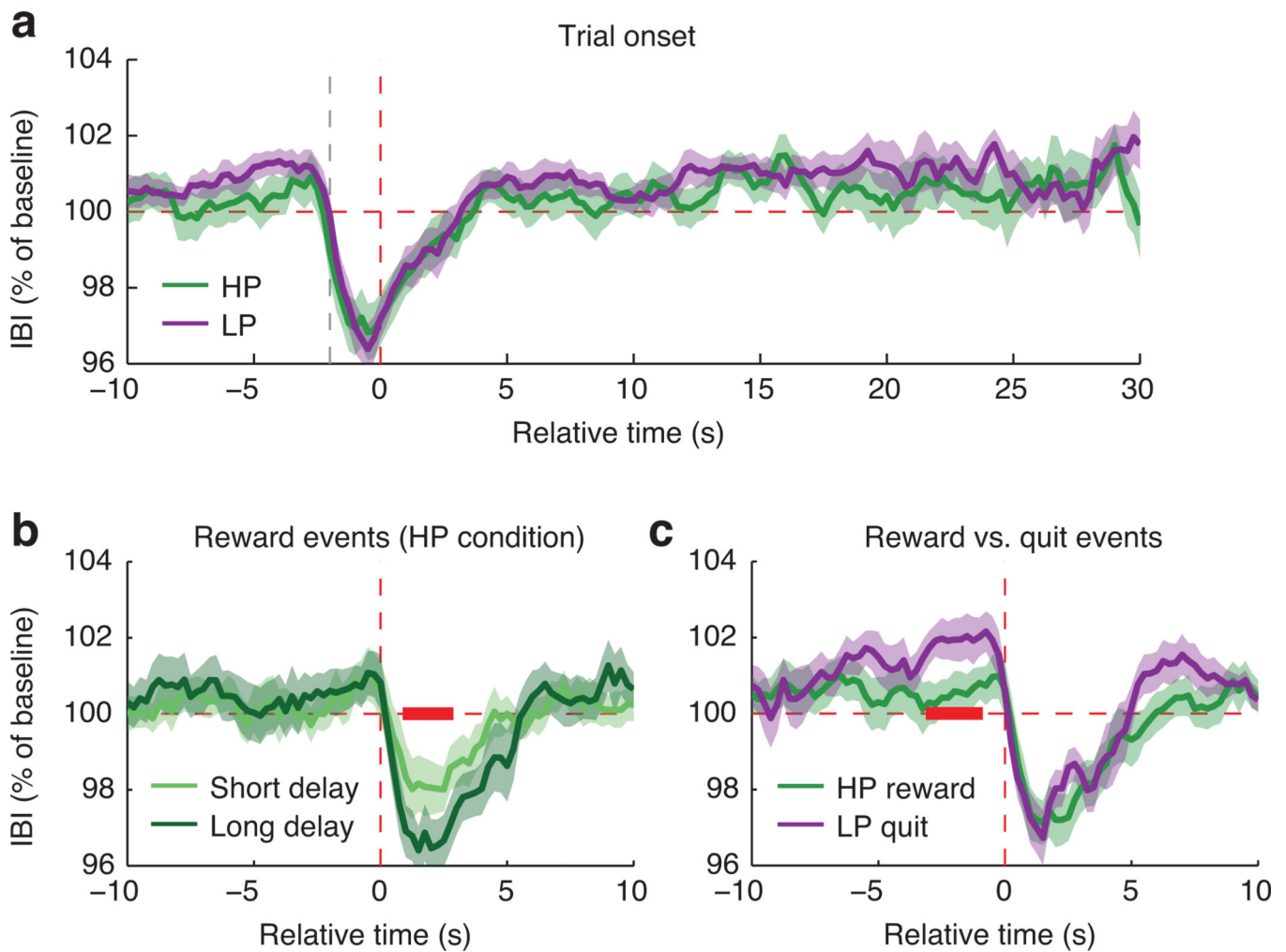
Figure 6.
Regions in which BOLD signal differentiated reward-related and quit-related keypresses, assessed on the basis of the event type (reward vs. quit) by timepoint interaction. Warm colors represent *F* statistics for the analysis of full timecourses, and crosshairs mark local peaks. Blue outlines mark regions significant in the analysis of pre-quit timepoints only. Timecourses (mean +/− SEM) are plotted for a 6mm-radius (33-voxel) sphere centered at each depicted focus point. Black dashed lines mark the keypress time; blue dashed lines mark the median reward cue time (for reward-related keypresses).

Figure 7.
Effects of task events on mean cardiac inter-beat interval (IBI; lower values correspond to faster heart rate). Error bands show SEM; red bands mark significant differences. A: Mean trial-onset-locked IBI timecourse in each condition. Vertical red dashed line marks trial onset; gray dashed line marks the preceding keypress. Each trial contributed data until 1 s before the trial ended (later timepoints therefore have fewer observations than earlier timepoints). No significant differences were observed. B: Comparison between rewards arriving at shorter (5–20s) vs. longer (25–40s) delays in the HP condition. The amplitude of post-keypress heart-rate acceleration was greater for rewards that followed longer delays (lag +1s to +2.75s; permutation-based $p=0.018$). C: Comparison between reward events in the HP condition and quit events in the LP condition, each restricted to trials with duration >10s. Vertical red dashed line marks the time of the reward cue or quit keypress. Results suggested transient cardiac deceleration prior to quit responses (lag −1s to −3s; permutation-based $p=0.045$).

Table 1

Region	x	y	z	Cluster extent	Peak value	Cluster *p* value
A: Trial-onset locked timecourses: Model-based contrast (t statistic)						
VMPFC	0	60	3	311	5.38	0.014
B: Trial-onset locked timecourses: Condition-by-time interaction (F statistic)						
L VMPFC	−3	60	3	84	5.54	0.001
R VMPFC	12	42	3	38	4.23	0.005
L posterior parietal	−42	−72	45	42	4.79	0.003
L superior temporal gyrus	−54	6	−15	16	4.09	0.034
C: Anticipatory activity in quit-related timecourses: Main effect of time (F statistic), −12.5s to −2.5s						
R posterior parietal	21	−72	51	393	9.32	0.001
L posterior parietal	−21	−72	48	59	5.20	0.042
R anterior insula	33	27	6	140	10.32	0.005
DMFC	3	12	45	127	7.45	0.007
R anterior PFC	33	54	27	70	7.65	0.024